



Les cafés de la statistique

"La statistique éclaire-t-elle les questions de société" ?

Soirée du 08 janvier 2013

Comparaisons internationales des niveaux d'éducation

Synthèse des débats ^[*]

Il y a un peu plus de dix ans, la diffusion des résultats de l'enquête « Pisa - Programme international pour le suivi des acquis des élèves » créait un choc en Allemagne. Persuadée de l'excellence de son système éducatif, l'Allemagne s'était retrouvée avec des résultats inférieurs à la moyenne des pays de l'OCDE ! Auparavant, la France avait mis en cause la méthodologie d'une enquête comparative sur la littératie des adultes.

Mais que nous apprennent au juste ces comparaisons entre pays ? Que veut-on mesurer précisément ? Comment les résultats sont-ils établis ? Un ensemble d'indicateurs chiffrés suffit-il pour comparer de façon pertinente les systèmes éducatifs, alors même qu'il fait nécessairement abstraction des contextes historiques et culturels dans lesquels chacun de ces systèmes s'est développé ? Quelles sont les limites de telles études et quels sont les effets (vertueux ou pervers) des publications très nombreuses qui en découlent ?

Invités :

Jean-Claude Emin, ancien secrétaire général du Haut conseil de l'évaluation de l'école¹ ;

Thierry Rocher, statisticien à la Direction de l'évaluation, de la prospective et de la performance [DEPP] du ministère de l'éducation nationale

Exposés introductifs

Pour introduire les débats, Thierry Rocher expose les principaux aspects de Pisa (qu'est-ce que c'est ? comment ça fonctionne ?), puis Jean-Claude Emin commente les résultats de cette enquête et les utilisations qui en sont faites.

Les principales caractéristiques de Pisa présentées par **Thierry Rocher** sont les suivantes :

^[*] Tant l'exposé liminaire que le contenu des échanges sont structurés en quelques thèmes, sans suivre l'ordre chronologique. Par ailleurs, l'identité des intervenants n'était pas toujours connue et l'on a choisi de ne pas attribuer nominativement les propos. Au reste, ceux-ci ont été reconstitués à partir des notes du secrétariat sans reprendre leur formulation détaillée. Pour retracer le débat, les thèmes sont souvent introduits sous forme d'une question : ce qui vient ensuite n'est pas la seule réponse de l'invité, mais l'ensemble des contributions des participants.

¹ Le **Haut Conseil de l'évaluation de l'école** a été remplacé en 2005, dans le cadre de la loi d'orientation et de programme pour l'avenir de l'école du 23 avril 2005, par le **Haut Conseil de l'éducation**, instance dotée d'une compétence consultative sur le socle des savoirs indispensables, les programmes et l'évaluation des résultats.

- il s'agit d'une évaluation internationale des compétences des élèves. Il existe des enquêtes internationales autres que Pisa ainsi que de multiples outils nationaux (notamment en France, pour évaluer le socle commun de connaissances et de compétences²), mais Pisa ambitionne d'évaluer le système éducatif dans son ensemble. Cette enquête triennale permet aussi depuis quelques temps des comparaisons diachroniques. Il faut avoir présent à l'esprit que sa dimension internationale est complémentaire des dispositifs d'observation nationaux qui, eux, poursuivent des objectifs propres à chaque pays ;
- des dispositifs d'évaluation internationaux ont existé dès la décennie 1950, sous l'égide de l'Unesco dont l'action fut relayée dans les années soixante par l'IEA³, association qui a promu depuis lors de nombreuses enquêtes de comparaisons internationales⁴. Pisa est l'héritière de ces travaux, aussi bien dans sa conception que dans ses méthodes. L'OCDE⁵, qui la met en œuvre, a beaucoup contribué à sa notoriété par ses capacités médiatiques ;
- par le soin apporté à sa conception et son déroulement, par l'ampleur des concertations entre statisticiens des pays concernés, Pisa est devenue une référence, un véritable standard méthodologique aujourd'hui, et influence fortement les dispositifs nationaux d'évaluation ;
- la cible de Pisa est la génération des élèves de 15 ans et non pas un niveau d'études comme, par exemple, dans les enquêtes TIMSS⁶. Pour la France, ces élèves sont en troisième ou en quatrième (pour ceux qui ont redoublé une ou deux fois) ou en seconde (pour ceux « à l'heure »), voire en 1^{ère} (pour les quelques-uns en avance), précision importante pour une bonne interprétation des résultats ;
- Pisa évalue les compétences, c'est-à-dire qu'on ne mesure pas les connaissances des élèves mais la capacité de ces derniers à les mobiliser dans un cadre plus large. Il ne s'agit pas non plus d'évaluer les programmes scolaires mais d'apprécier l'aptitude des élèves à mobiliser des acquis (notamment scolaires) hors du cadre scolaire. Cette idée est d'ailleurs reprise en France dans la définition du socle commun de connaissances et de compétences⁷ ;

² Cf. décret n° 2006-830 du 11 juillet 2006. Il s'agit des connaissances et compétences attendues des élèves en fin de scolarité obligatoire.

³ International Association for the Evaluation of Educational Achievement

⁴ Voir en annexe 1 les planches 7 et 8 du diaporama.

⁵ L'Organisation de coopération et de développement économique (OCDE) a, en 1961, succédé à l'Organisation européenne de coopération économique (OECE), fondée en 1948 pour gérer l'aide américaine d'après-guerre (plan Marshall). L'OCDE regroupe plusieurs dizaines de pays : toute l'Europe occidentale et l'Amérique du nord, le Japon, l'Australie, la Nouvelle-Zélande, la Corée et, depuis quelques années, certains pays d'Europe centrale (République tchèque, Hongrie, Pologne) et d'Amérique latine. L'OCDE est le principal rassembleur de statistiques sur les pays développés. Elle siège à Paris.

⁶ TIMSS pour Trends in International Mathematics and Science Study

⁷ Extrait de l'annexe du décret mentionné en note 2 : « ...le socle commun est le ciment de la nation : il s'agit d'un ensemble de valeurs, de savoirs, de langages et de pratiques dont l'acquisition repose sur la mobilisation de l'école et qui suppose, de la part des élèves, des efforts et de la persévérance. La définition du socle commun prend également appui sur la proposition de recommandation du Parlement européen et du Conseil de l'Union européenne en matière de compétences clés pour l'éducation et l'apprentissage tout au long de la vie. Elle se réfère enfin aux évaluations internationales, notamment au Programme international pour le suivi des acquis des élèves (PISA), qui propose une mesure comparée des connaissances et des compétences nécessaires tout au long de la vie. ». On remarquera la référence explicite à Pisa.

- Pisa évalue trois domaines :
 - la compréhension de textes écrits ; on parle ici plus précisément de littératie⁸ ;
 - la culture mathématique ; on trouvera page 11 de l'annexe 1 un exemple de question (on dit « d'item ») de Pisa sous cette rubrique ;
 - la culture scientifique ;
- ce que l'on mesure est spécifique à Pisa ; on considère que les compétences ne sont pas observables directement mais que leurs manifestations le sont. Si bien que la compétence est en quelque sorte ce que mesure Pisa, de même qu'Alfred Binet⁹ définissait l'intelligence comme étant ce que mesuraient ses tests. L'instrument de mesure et l'objet observé sont indissociables et cela impose un grand luxe de précautions pour assurer la validité des comparaisons internationales : il faut veiller à la pertinence, la précision et l'unidimensionnalité de la mesure, celle-ci devant permettre un classement univoque des élèves indépendamment des cultures. Cela signifie que si, de deux items A et B, le premier est mieux résolu que le second en France, il doit en être de même dans tout autre pays ; sinon, c'est qu'il existe un fonctionnement différentiel de Pisa d'un pays à l'autre pour cet item. Toutes sortes de facteurs peuvent intervenir dans cette mise au point des items : ainsi de l'utilisation ou non de supports textuels, ou bien de questionnaires à choix multiples (QCM) ou de questions appelant la rédaction d'une réponse¹⁰. Le constat éventuel de la spécificité culturelle d'une réponse est intéressant par lui-même, puisqu'il justifie des investigations de nature à permettre l'amélioration du questionnement ;
- l'enquête se fait sous l'égide de l'OCDE avec la participation de pays hors OCDE. Elle est collaborative au sens où on recherche le consensus entre les pays sur les questions posées¹¹ ;
- l'échantillon compte en France 4 500 à 5 000 élèves et donne des résultats nationaux. Au Canada, les autorités veulent obtenir aussi des résultats pour chaque province. Pisa 2009 a donné lieu à des questionnaires traduits en 44 langues selon des procédures très strictes. L'importance des problèmes de traduction a été mise en évidence par des travaux de l'Ined¹² ; en 1990, pour des questions de formulation, la France s'est retirée d'une enquête internationale sur les adultes. Pisa en a tiré les enseignements ;
- le déroulement de l'enquête dans l'ensemble des pays participants est très encadré, beaucoup plus que dans la plupart des évaluations analogues, avec de multiples contrôles de qualité et de nombreux standards - par exemple en matière de non-réponse - et des possibilités de sanction si un standard n'est pas respecté. Ainsi, un pays peut être exclu de la publication de ses résultats¹³. C'est arrivé en Grande-Bretagne et aux Pays-Bas ;

⁸ La notion de « **littératie** » recouvre les fondamentaux que le dispositif éducatif est censé inculquer pour adapter l'élève à une société de culture écrite.

⁹ Pédagogue et psychologue français connu pour ses travaux sur la psychométrie.

¹⁰ Par exemple, aux États-Unis, les élèves répondent davantage que les élèves français quand ils ont à rédiger. En outre, contrairement à certaines idées reçues, les élèves français ont globalement de meilleures performances aux questions à choix multiples.

¹¹ Certains observent avec malice que ce consensus est certes trouvé, mais dans le giron du secrétariat de l'OCDE...

¹² Institut national d'études démographiques

¹³ Ce qui peut l'arranger ! pensent certains participants.

- les élèves passent les tests sous forme papier-crayon pendant deux heures en répondant à des questions ouvertes et des questions fermées. En 2015, on ira peut-être vers une version informatisée de Pisa, ce qui posera – soit dit en passant - des problèmes de comparabilité dans le temps. Quelque forme que prennent les tests, la question de la motivation des élèves reste posée, autant que celle des causes de la non-réponse.

On pourrait s'étonner d'un tel luxe de précautions dans la conception et la réalisation de Pisa. Cela résulte des multiples travaux effectués en matière de docimologie¹⁴ et du constat des variations dans les corrections d'une même copie de concours ou d'examen. Pisa a donc un système de quadruple correction ! Il subsiste néanmoins des problèmes de fond ; il convient par exemple de veiller à éviter toute triche car il y a évidemment un aspect de compétition internationale dans la diffusion des résultats de Pisa. À noter qu'il n'y a pas en France d'incitation financière consentie aux établissements d'enseignement participants, ni aux élèves participants.

Jean-Claude Emin précise que Pisa a été exécutée en 2000, 2003, 2006, 2009 et 2012, avec chaque fois, par rotation triennale, l'un des trois domaines d'évaluation mis en majeur. La dernière enquête (2012) est en cours de dépouillement et les résultats seront publiés fin 2013. La comparaison des enquêtes Pisa de 2000 et 2009 est particulièrement intéressante puisque ces deux années-là le domaine de la compréhension de l'écrit était mis en majeur et qu'on dispose des scores des 26 pays communs aux deux enquêtes. On en tire les quelques constats qui suivent en ce qui concerne notamment la littératie (définie comme il a été dit plus haut) :

- en France, les élèves de 15 ans sont dans la moyenne de l'OCDE (c'est vrai dans les trois domaines couverts). Les pays géographiquement proches de la France sont proches aussi en résultats. La Finlande et la Corée sont en tête du classement en points (la Corée se distinguant pour les sciences), étant précisé que la moyenne n'a aucune valeur normative ;

- on constate une dispersion importante, plus importante qu'ailleurs, de nos élèves dans ce domaine de la compréhension de l'écrit. Cela doit nous interpeller sur notre système éducatif et sur le redoublement¹⁵. On constate en effet qu'il y a un écart-type d'éloignement entre les élèves de 15 ans qui sont en seconde et les élèves de 15 ans qui sont en troisième ou en quatrième ;

- en évolution, on constate une baisse légère (confirmée par différentes autres évaluations extérieures à Pisa) du niveau moyen de compétence des élèves français, surtout due à l'augmentation du nombre des élèves les moins performants aux tests. Autrement dit, notre système éducatif laisserait de plus en plus d'élèves sur le bord du chemin et dans un état de moins en moins satisfaisant en termes d'adaptation à l'écrit, sans qu'on constate beaucoup de mieux à l'autre extrémité du spectre ;

- c'est en France que les résultats semblent le plus fortement liés au statut socioprofessionnel des familles des élèves. Sur ce critère, la France est le deuxième plus mauvais pays après la Nouvelle-Zélande. Ce constat frappe surtout les garçons et traduit une certaine iniquité sociale.

¹⁴ Etude de la qualité et de la validité des différents systèmes de notation scolaire et de contrôle des connaissances.

¹⁵ On redouble plus en France qu'en Europe. Nos élèves de 15 ans qui ne sont pas encore en seconde n'ont pas tous redoublé. Parmi ceux qui ont redoublé, la moitié environ l'ont fait dans le primaire et la moitié au collège.

Que conclure en nous comparant à d'autres systèmes éducatifs ? Certes, la médiatisation de Pisa conduit à des palmarès qui détournent parfois l'attention des véritables problèmes mais on voit tout de même que dans les systèmes les plus performants en termes de points obtenus (systèmes qui sont souvent les plus équitables), on rencontre fréquemment certaines des caractéristiques suivantes :

- les pays considérés pratiquent peu de différenciations : pas d'orientation précoce des élèves, mixité sociale, peu ou pas de redoublements, pas d'éviction des écoles des élèves difficiles ou moins performants ;
- ils accordent aux écoles un degré d'autonomie plutôt important en matière de curriculum, associé à une faible compétition entre écoles quant au recrutement des élèves (autrement dit la liberté de choix de l'école apparaît comme un facteur plutôt négatif) ;
- ils ont fait le choix d'avoir des enseignants bien payés plutôt que celui de faibles effectifs dans les classes.

Bien sûr, tout cela appelle beaucoup de nuances et il existe des pays performants qui ne présentent pas tous ces caractères et d'autres qui ont des traits de cette nature sans briller par leur résultat d'ensemble.

De tels constats ne passent pas inaperçus ; d'autant moins que, par sa médiatisation, Pisa a donné indirectement davantage d'écho à d'autres évaluations pratiquées par ailleurs. Le discours politique est passé en une dizaine d'années de l'indifférence à la référence obligée à Pisa. L'indifférence était de mise quand les résultats de la France étaient plutôt bons ; leur relative dégradation et certaines spécificités françaises ont appelé l'attention. Par exemple, le taux élevé de non-réponse en France pose beaucoup de questions ; on peut avancer l'idée que l'erreur est considérée chez nous davantage comme une faute que comme un moyen de progresser, ce qui inciterait les élèves passant les tests Pisa à ne pas répondre s'ils craignent de donner une mauvaise réponse. Cela rejoindrait l'idée que les élèves français sont assez stressés et ont tendance, plus qu'ailleurs, à se dévaloriser. La rédaction des réponses fournies aux tests en France révèle aussi des problèmes avec l'écriture.

Le personnel politique est donc devenu attentif, mais il est significatif de constater que le dernier ministre de l'éducation nationale sous la présidence de Nicolas Sarkozy (Luc Chatel) a rendu compte avec beaucoup d'objectivité des résultats de l'enquête Pisa 2009 mais est resté muet sur les conclusions qu'on devait en tirer pour l'évolution du système éducatif du pays. Pour sa part, le ministre actuel (Vincent Peillon) veut donner la priorité au primaire, et le justifie en s'appuyant notamment sur les résultats d'une récente enquête, l'enquête PIRLS 2011 (Progress in International Reading Literacy Study), qui mesure les performances en lecture des élèves de CM1¹⁶

Les enseignants sont peu au courant pour leur part des résultats de Pisa mais les responsables des organisations syndicales y sont plus attentifs et ont contribué à mieux faire comprendre les évaluations et l'utilisation qu'on peut en faire. Ainsi, les enseignants sont peu à peu sensibilisés aux évaluations dans l'ensemble des classes.

¹⁶ Note d'information n°12-21 du ministère de l'éducation – décembre 2012

Débat

Un peu de technique statistique

En réponse à une question sur la façon dont les établissements et les élèves sont choisis dans Pisa, il est précisé qu'un échantillonnage aléatoire rigoureux est opéré au niveau des établissements, puis des élèves, dans le respect de la théorie des sondages. Les établissements incluent les établissements de l'enseignement privé et ceux des départements d'outre-mer (à l'exception de La Réunion à cause des calendriers scolaires). L'ensemble de la population de 15 ans scolarisée est couvert, à l'exception des handicapés, qui posent un problème spécifique.

On établit le score atteint par chaque élève en réponse aux différents items, qui sont d'inégale difficulté, on construit une échelle des scores (de moyenne 500 et d'écart-type 100¹⁷), puis on définit des tranches de score qui déterminent des niveaux de compétence. Un élève est affecté à un niveau en fonction de son score, et cela correspond à une probabilité de réussite aux items de ce niveau au moins égale à 50 %. Autrement dit, la définition des niveaux de compétence se fait à partir des résultats constatés et des libellés des items et non pas a priori ; cette précision pour répondre à un participant qui se demandait comment on parvenait à discriminer par des items entre deux niveaux comme les 5 et 6¹⁸, dont les différences lui paraissaient d'une subtilité extrême.

On trouve dans le rapport de l'OCDE sur Pisa 2009 quelques exemples d'items mais tous les items ne sont pas accessibles (ne sont pas « libérés ») afin d'éviter les risques de préparations des élèves organisées par les établissements en amont de la passation des tests.

Le graphique de la page trois du document auquel il est renvoyé par un lien en fin de compte rendu (document constituant l'annexe 3) laisse perplexe une participante pour ce qui concerne la colonne de droite, qui met en évidence une dispersion considérable des sous-groupes d'élèves français selon qu'ils sont « à l'heure » ou en retard dans leur cursus scolaire. Qu'en est-il dans les autres pays ? Certes, comparer des moyennes de sous-groupes d'élèves français à des moyennes de pays est peu orthodoxe mais l'exercice visait seulement à discerner, à âge égal, l'influence du retard dans le cursus scolaire et à comparer cet écart à celui observé – avec les mêmes items - entre les niveaux de compétence moyens d'élèves de 15 ans et d'élèves de 14 ans, tous « à l'heure »¹⁹. Ce dernier étant bien moindre que le précédent, on est conduit à beaucoup d'interrogations sur les vertus du redoublement²⁰. Il semble que les performances de sous-groupes analogues soient plus resserrées en

¹⁷ Echelle construite lors de Pisa 2000. Elle a pu se déformer depuis lors.

¹⁸ Niveau de compétence 6 : être capable de comprendre de façon fine et détaillée un ou plusieurs documents, d'émettre des hypothèses d'analyse ou de juger d'un regard critique un texte complexe portant sur un thème peu familier.

Niveau 5 : être capable d'évaluer l'information et d'élaborer des hypothèses, en faisant appel à des connaissances spécialisées, en développant des concepts contraires aux attentes.

¹⁹ Pour les besoins de la cause, un échantillon d'élèves de 3^e « à l'heure » avait été constitué.

²⁰ L'écart entre élèves de 3^e de 15 ans en retard et élèves de 2nde de 15 « à l'heure » cumule deux « effets » : retard scolaire et effet d'apprentissage d'une année scolaire. C'est pourquoi on a également évalué les élèves de 3^e de 14 ans à l'heure. Si on avait constaté un écart très important entre les 3^e de 14 ans à l'heure et les 2nde de 15 ans à l'heure, l'interprétation de

Finlande qu'en France mais il y aurait par ailleurs en Finlande une plus grande inégalité pour le passage à l'enseignement supérieur selon les milieux socio-culturels.

Un participant s'interroge sur les indications que PISA est susceptible de fournir à propos de catégories d'établissements, mais cette question n'est pas abordée par Pisa en France en raison du choix de l'âge de 15 ans. Il se trouve que dans notre pays tous les élèves de 15 ans qui n'ont eu aucun accident de parcours sont au lycée et tous les autres élèves scolarisés de la même génération au collège. Si on comparait les résultats moyens de Pisa par catégorie d'établissements, on comparerait en fait des élèves en retard (collèges) et des élèves « à l'heure » (lycées). On aurait une variance entre établissements de 50 % avec ce type d'approche alors qu'elle est ailleurs de 15 à 25 %. En revanche si on avait fait passer Pisa à 14 ans, on aurait eu presque tous les élèves au collège. Le choix de la limite est résulté d'un compromis international, l'âge de 15 ans étant apparu le moins mauvais compromis pour la fin de la scolarité obligatoire. Lors de Pisa 2000, la France a montré le problème à l'OCDE ; comme cette organisation ne veut pas publier des tableaux à trous, certaines questions destinées aux chefs d'établissement ont été supprimées. Mais un lobbying interne à l'OCDE va redonner vie à ces questions et les résultats par catégorie d'établissement qui seront publiés n'auront guère de signification pour la France.

Un autre problème serait de faire des comparaisons entre les niveaux des élèves à différents stades de leur vie scolaire : comment le système éducatif fait-il progresser les élèves ? On a là une notion de valeur ajoutée d'un établissement entre l'entrée et la sortie de l'élève. Des travaux existent à ce sujet, mais pas dans le cadre de Pisa.

Suspensions ?

Un participant s'interroge sur l'idéologie dominante au secrétariat de l'OCDE car il nourrit un certain scepticisme à l'égard des choix des items de Pisa et de la pertinence de l'instrument. Ainsi, les élèves français obtiendraient dans le domaine de la culture mathématique, sur des items touchant des sujets ne figurant pas au programme de leurs cours, des résultats analogues à ceux des élèves d'autres pays ! Ou encore, la Société mathématique de France aurait salué le niveau de compétence des élèves de quinze ans en Finlande mais, de l'aveu même des Finlandais, le constat serait beaucoup moins optimiste pour leurs étudiants de dix-huit ans par rapport à des étudiants du programme Érasmus²¹. Pourtant, la publicité faite à Pisa est susceptible d'entraîner des décisions lourdes de conséquences. Au sortir de Pisa 2000, l'Allemagne a éprouvé un véritable choc, elle qui croyait avoir le meilleur système éducatif du monde ; n'est-elle pas allée en conséquence jusqu'à supprimer la pratique du sport tous les jours à 14 h. ?

Il y aurait sans doute beaucoup à dire sur l'idéologie mais on ne peut imaginer qu'une idéologie dominante soit consciemment mise en œuvre par les multiples acteurs des dizaines de pays

la différence entre les deux groupes en retard/à l'heure aurait été différente. Quoi qu'il en soit, on ne peut pas, en toute rigueur, parler ici d'une inférence directe du redoublement mais d'un indice supplémentaire parmi un faisceau de résultats concordants par ailleurs.

²¹ Erasmus (pour *European Region Action Scheme for the Mobility of University Students*) est le nom donné au programme d'échange d'étudiants et d'enseignants entre les universités et les grandes écoles européennes. C'est un sous-ensemble du programme Éducation et formation tout au long de la vie (EFTLV). Le nom du programme vient de celui du moine humaniste et théologien néerlandais Érasme (1469 ?-1536).

concernés. Chacun raisonne avec ses propres références culturelles et techniques et l'expérience montre qu'il est plus ardu de parvenir au consensus pour les mathématiques et les sciences que pour la littérature. Cela tient avant tout au fait que les programmes scolaires sont très différents selon les pays ; on en voit trace dans la complexité des enquêtes TIMSS. Par ailleurs, les compétences mesurées par Pisa peuvent trouver leur source aussi bien dans le milieu extra scolaire que dans la classe. Et ce que révèle Pisa est un objet de réflexion, voire d'action, pour chaque pays.

Pourquoi, demande une participante, l'OCDE a-t-elle évincé l'Unesco de ce domaine qui aurait dû rester le sien sous l'égide de l'ONU ? Ainsi l'Unesco depuis 1950 avait lancé de multiples travaux avec des enquêtes sur le fonctionnement de l'éducation et une recherche de définitions précises : qu'appelle-t-on un élève ? qu'est-ce qu'un nombre d'élèves ? etc. En fait, l'OCDE a été sollicitée d'intervenir par les USA, rejoints par la France et l'Allemagne. Il y avait dans le monde libre des besoins de données pour améliorer les indicateurs et l'OCDE, organisme intergouvernemental, paraissait mieux à même de coordonner une démarche d'ensemble, indépendamment de toute idéologie. Aujourd'hui, la moitié de son chiffre d'affaires est réalisée avec les résultats de Pisa ! La direction de l'éducation de l'OCDE a même été créée dans cette mouvance.

A la question de savoir qui prend les décisions relatives au contenu de Pisa, la réponse est que les statisticiens des différents pays sont au premier rang des discussions²², sans qu'on sache très bien, pays par pays, s'ils défendent des positions purement techniques ou si le politique a voix au chapitre en arrière-plan. Il y a bien sûr des enjeux politiques aux résultats : dans des pays comme les Etats-Unis ou la Grande Bretagne, ces enjeux sont énormes, et pour les établissements et pour les enseignants. Dès lors, voyant des pays progresser fortement d'une enquête à l'autre, certains posent la question d'un biais éventuel pour forcer les résultats à monter ; ou, plus banalement, d'un effort d'entraînement systématique des élèves pour qu'ils passent mieux l'épreuve ; ou encore de formes d'enseignement plus proches, dans ces pays, de l'esprit des tests Pisa. Mais l'objectif d'un système éducatif doit-il être de réussir à Pisa ? La réponse est assurément négative en France et ce point est important.

Limites des comparaisons

Les performances de la Corée (du Sud) et de la Finlande sont source de perplexité pour une participante : voilà deux pays qui sont en tête et qui n'ont pas du tout le même système éducatif. Que penser de cela ? Tout ce qu'on peut dire est qu'avec ce que mesure Pisa, les élèves scolarisés à 15 ans ont les mêmes résultats dans les deux pays. Après, si on regarde les deux sociétés et les deux systèmes éducatifs en cause, on peut se poser mille questions et faire des hypothèses. En tout cas la Finlande est très visitée pour son système éducatif et on peut noter que l'insertion de l'école dans la société (et vice versa) y est très forte.

À moyens égaux, les systèmes éducatifs réussissent mieux dans certains cas que dans d'autres. Des facteurs politiques et sociétaux sont à l'origine de ce fait, qui est en soi un objet de recherche.

²² Claude Thélot, ancien directeur de la DEPP, regrettait que l'OCDE n'ait pas davantage d'experts français.

Un participant, tout en se déclarant d'accord sur les concepts de Pisa et sa méthodologie, relève que les comparaisons internationales peuvent apparaître aussi délicates que les comparaisons entre régions en France dès lors que les palmarès qu'on en tire n'analysent pas la nature et l'intensité de l'investissement consenti par les milieux politique et économique et par les familles dans les systèmes éducatifs ainsi évalués. Une autre personne souligne à quel point la confiance dans les systèmes éducatifs est inégale selon les lieux et met cela en relation avec l'iniquité ressentie de ces systèmes lorsqu'ils n'accordent pas de moyens différenciés en fonction du degré de réussite ou de non réussite des élèves.

Il n'en reste pas moins que Pisa (et d'autres outils) apportent une mesure de l'efficacité des systèmes éducatifs, certains étant à l'évidence confortés par un investissement important consenti par le corps social. Même la manière dont l'école est considérée par la société peut avoir des effets sur son efficacité. Et, comme cela a été déjà souligné, Pisa ne mesure pas que les effets de la vie scolaire mais aussi ceux de la vie courante, qui apporte d'autres acquis.

Un autre participant observe que les résultats de Pisa reflètent aussi les différences d'homogénéité entre pays ; ainsi, il y a pas ou peu d'immigrés en Finlande, ce qui peut être mis en relation avec les performances de ce pays dans Pisa. Cette remarque déclenche plusieurs réactions. Pour les uns, la réponse appropriée au constat de compétences moindres des populations scolaires issues de familles immigrées est celle du Canada, qui donne pour objectif à son école d'homogénéiser les compétences. D'autres mentionnent que toutes les études conduisent à la conclusion que, à même catégorie socioprofessionnelle du milieu familial, les performances des allochtones sont proches de celles des autochtones²³. D'autres encore admettent le constat d'une grande variance des résultats en France, d'une dérive vers le bas et d'une moindre correction des inégalités par notre système éducatif mais rejettent toute passivité politique devant ce constat. De ce point de vue, n'est-il pas inquiétant d'entendre que les systèmes performants sont plutôt indifférenciés alors qu'on attendrait des mesures spécifiques en direction des élèves moins performants ? En réalité, si on parle d'enseignements différenciés en France, c'est en référence aux filières d'enseignement, dont le fonctionnement paraît en fin de compte moins performant que le système des classes homogènes dans d'autres pays.

Il est clair en tout cas que l'OCDE livre des résultats qui sont de l'ordre du constat et ne comportent aucune approche normative ni de jugement idéologique concernant les allochtones.

Evaluer l'évaluation ?

Le système est-il soumis à l'épreuve de la recherche, demande une participante ? Par exemple, étudie-t-on le biais éventuel introduit dans Pisa par la formulation des items ? Prenons l'item présenté page 11 de l'annexe 1 : il relève de la pure culture mathématique puisqu'il s'agit de déterminer l'un des deux termes d'un quotient connaissant l'autre terme et le résultat. Or, ayant soumis l'item en question à son entourage, la participante a constaté que la plupart des gens voulaient comprendre le problème physiquement et, de ce fait, le trouvaient difficile alors qu'elle-même l'avait résolu mécaniquement sans difficulté. Est-ce neutre ? De la même manière, les

²³ Selon un participant, ce serait inexact s'agissant des élèves d'origine turque.

constantes observées dans les pays performants à Pisa sont certes intéressantes mais le constat est empirique ; y a-t-il des recherches qualitatives et quantitatives sur le sujet ?

Pisa est assurément un objet de recherche, et toutes les observations sont accessibles en ligne sur Internet. Mais force est de constater que la tradition psychométrique est faible en France où on est plus porté vers la philosophie cognitive. Il n'y a pas assez de recherches sur ce que l'on mesure à travers Pisa. Ou sur la neutralité des items. Ou encore sur ce qui motive la non-réponse : la peur ? l'incompréhension ? le stress (comme le pensent les Suédois de nos élèves) ? Quand on parle avec les élèves, on voit que les stratégies sont très différentes d'une personne à l'autre. Et puis, l'enveloppe de crédits pour les travaux de recherche est limitée.

Pourtant, la recherche aide à se poser les bonnes questions, notamment sur les objectifs du système éducatif. Le mérite de Pisa est de conduire à cela, avec ou sans recherche pointue. De ce point de vue, sur le plan politique, le forcing de l'OCDE pour faire entrer dans Pisa des pays qui n'ont à leur disposition que les comparaisons internationales peut les conduire à mieux réfléchir à leurs objectifs propres.

Et demain ?

Comment l'enquête Pisa évolue-t-elle ? Assurément par petites touches, pour trois catégories de raisons :

- les comparaisons dans le temps appellent une certaine stabilité de l'instrument de mesure ;
- il y a des résistances institutionnelles au changement. Quatre-vingts pays sont maintenant concernés, dont il faut recueillir l'accord ! Et cela indépendamment du travail technique d'harmonisation des items chaque fois qu'un nouveau contexte culturel est à prendre en compte. En outre, l'OCDE fonctionne par appels d'offres pour la mise en œuvre de Pisa et constitue des consortiums internationaux pour chaque enquête²⁴. Le fait qu'à deux reprises le consortium ait été mené par un même pays (en l'occurrence, l'Australie) a pu favoriser un certain conservatisme ;
- enfin, il y a les aspects financiers. Pisa coûte cher : 230 000 € sont versés par la France à l'OCDE et les coûts internes à l'éducation nationale avoisinent la même somme. La France doit en outre financer ses dispositifs nationaux d'évaluation.

Un participant objecte qu'un coût de 200 000 à 300 000 € n'est qu'une goutte d'eau dans le budget de l'éducation nationale, à quoi il est répondu que le coût est faible en effet dans l'absolu mais que les enveloppes budgétaires pour les enquêtes et la recherche ne le sont pas moins !

Les choses évoluent cependant. On a parlé plus haut de l'expérimentation de la lecture sur support électronique dans Pisa 2009²⁵. Pisa 2015, qui aura le domaine scientifique en majeur, devrait introduire des items sur l'informatique. Peut-on espérer, demande une participante, une évaluation par Pisa des compétences transversales aux trois domaines qu'elle couvre ? C'est peu probable avant pas mal de temps car il s'agirait d'un stade plus élaboré et plus complexe encore de la mécanique Pisa. Avant cela, on envisage d'évaluer l'apport des systèmes éducatifs dans le cursus des élèves : il faudrait constituer pour cela des cohortes (des panels dans le vocabulaire de Pisa) d'élèves

²⁴ La Direction de l'évaluation, de la prospective et de la performance (DEPP) du ministère de l'éducation est l'opérateur de Pisa en France et a participé à deux consortiums.

²⁵ Cette expérimentation (Electronic Reading Assessment) a été menée dans vingt pays, dont la France.

de la sixième à la troisième afin de mesurer l'efficacité du collège. Une réflexion est amorcée dans Pisa mais c'est un projet extrêmement lourd et extrêmement complexe.

Dans les évolutions à venir, il ne faut pas négliger que de plus en plus de pays sont concernés par Pisa. Jusqu'ici les entrants devaient adopter les concepts de Pisa, qui sont à vrai dire plus anglo-saxons que latins. Les choses bougent avec l'entrée de nouveaux pays – latino-américains et est-européens, notamment - dans l'OCDE elle-même ; la Russie frappe à son tour à la porte. Le consensus devra probablement évoluer, de ce fait, vers un nouveau compromis sur la question du bagage intellectuel souhaitable pour un élève de 15 ans dans le monde du 21^e siècle.



Annexes :

- Annexe 1 : Panorama des évaluations (diaporama) ;
- Annexe 2 : Les enquêtes internationales en France : de l'indifférence à la « référence » - Jean-Claude Emin :
- Annexe 3 : Note d'information 10-24 de la DEPP (décembre 2010) consacrée aux résultats de l'enquête "PISA2009" ; voir :
http://media.education.gouv.fr/file/2010/99/8/NIMENJVA1024_161998.pdf



M@dos

Thématique : les évaluations internationales
des acquis des élèves

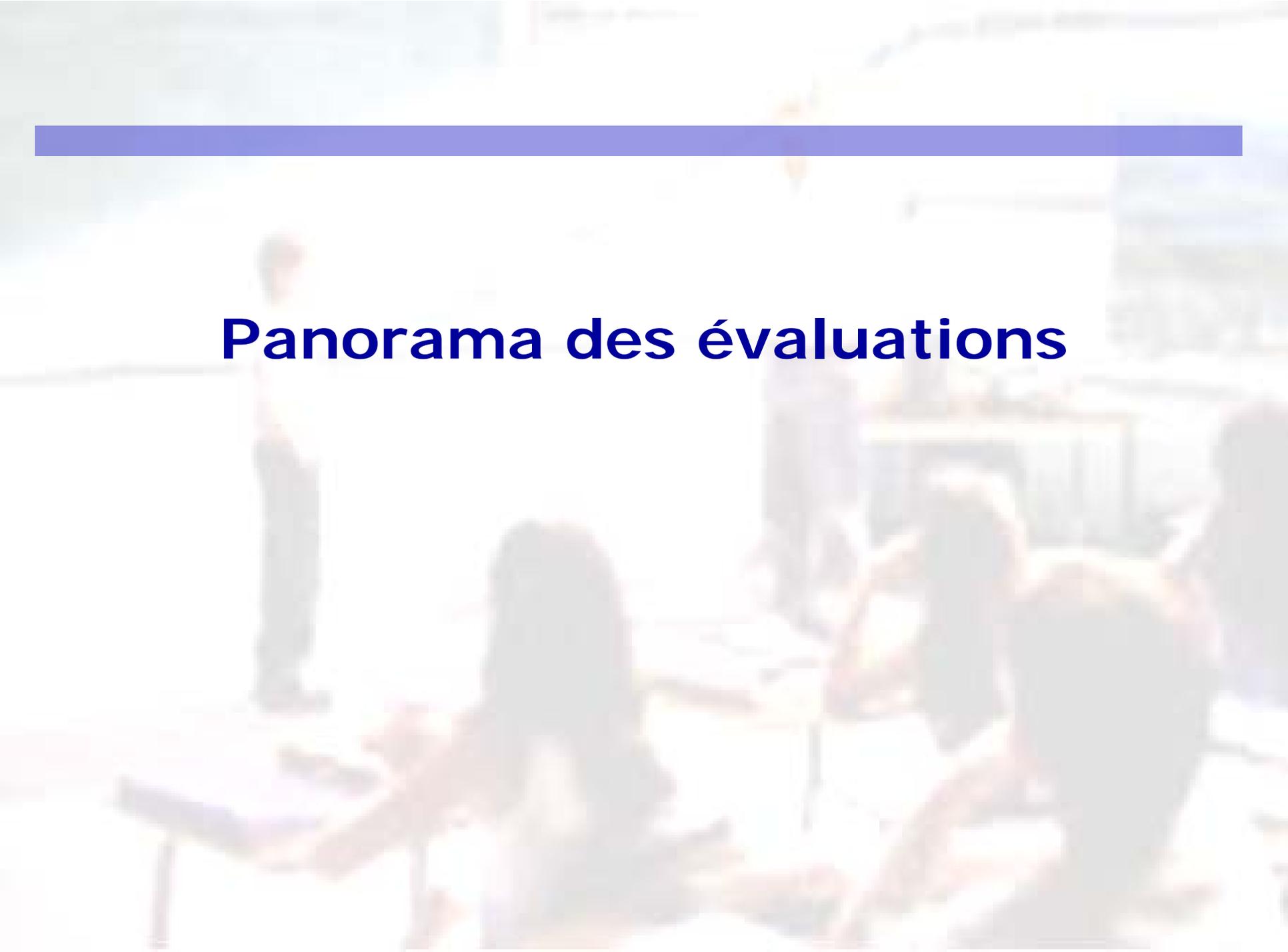
Présentation générale

Thierry Rocher

Classe centra, 28 novembre 2012

Plan

- Panorama des évaluations
- Cadre général
- PISA : organisation et méthodes
- PISA 2009

A blurred background image of a classroom. A teacher is standing at the front, and several students are seated at desks, some looking towards the front. The image is intentionally out of focus to emphasize the text overlay.

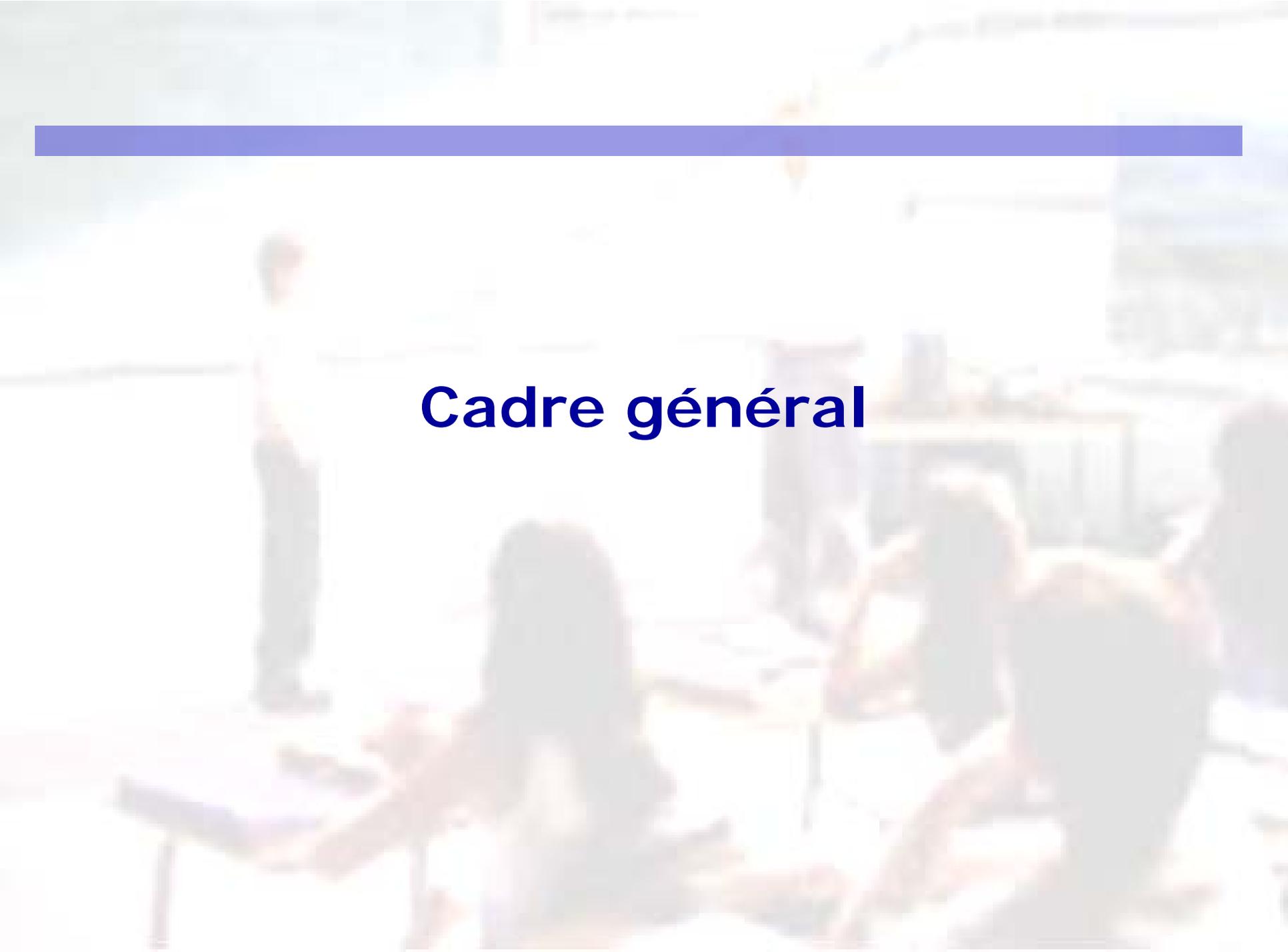
Panorama des évaluations

Les dispositifs d'évaluations (1)

- **CEDRE :**
 - ✓ cycle des évaluations disciplinaires réalisées sur échantillons en CM2 et en 3^{ème}, une évaluation spécifique chaque année, reprise tous les six ans
- **Indicateurs annuels :**
 - ✓ Indicateurs LOLF : maîtrise des compétences de base en français et en mathématiques en CM2 et en 3^{ème} (2007-2012), maîtrise des compétences du socle (depuis 2011-en construction)
 - ✓ Les jeunes de 17 ans en difficulté de lecture (JDC, ex-JAPD)
 - ✓ Évaluations exhaustives CE1/CM2 (et 5^e)
- **Comparaisons longitudinales :**
 - ✓ Suivi d'une cohorte d'élèves entrés en 6^{ème} en 2007
 - ✓ Suivi d'une cohorte d'élèves entrés au CP en 2011

Les dispositifs d'évaluations (2)

- **Comparaisons « historiques » :**
 - ✓ Compétences en lecture et en calcul des élèves de CM2 (1987-2007)
 - ✓ Difficultés de lecture à l'entrée en 6^{ème} (1997-2007)
- **Comparaisons internationales :**
 - ✓ PIRLS, PISA, Learning to learn, compétences en langues (ESLC)
- **Evaluations d'expérimentations :**
 - ✓ EIST, Lecture en GS, livrets de compétences, ...
- **Chez l'adulte** (littératie et numératie) :
 - ✓ IVQ (sous l'égide de l'INSEE : 2004, reprise en 2011)
 - ✓ PIAAC (sous l'égide de l'OCDE : 2011)

A blurred background image of a classroom. A teacher in a white shirt is standing on the left, pointing towards a whiteboard. Several students are seated at desks, some looking towards the teacher. The scene is brightly lit, and the overall image is out of focus.

Cadre général

Objectifs, utilisations

- Une vieille idée (années 50, UNESCO) :
 - ✓ « Le monde comme laboratoire »
 - ✓ La comparaison comme révélateur
 - ✓ Une compétition internationale
- Réception / utilisation :
 - ✓ Réticences : mesure-t-on la même chose dans tous les pays ? Et si oui de manière équitable ?
 - ✓ Idées reçues : méconnaissance du domaine en France (ex. : les QCM)
 - ✓ Objectif politique : « améliorer le classement de la France aux évaluations internationales »
 - ✓ Médiatisation et instrumentalisation des résultats (cf. rapport Know & Pol)

Histoire(s)

- Les principaux acteurs :
 - ✓ IEA (International Association for the Evaluation of Educational Achievement) : depuis les années 60, FIMS (1964), SIMS (1977), Reading Literacy (1991), TIMSS (1995, 1999, 2003, 2007, 2011), PIRLS (Progress in International Reading Literacy Study, 2001, 2006, 2011)
 - ✓ OCDE : PISA (Programme International pour le Suivi des Acquis des élèves, 2000, 2003, 2006, 2009, 2012)
 - ✓ Union Européenne : enquêtes ponctuelles, ESLC (European Survey of Language Competences, 2011)
 - ✓ CONFEMEN (PASEC depuis 1991)
- La référence aujourd'hui : PISA avec environ 450000 élèves de 15 ans dans 65 pays (2009)

Cadre de PISA

- PISA :
 - ✓ héritier de 50 ans d'expériences (organisation, méthodologie, publications ...)
 - ✓ se distingue sur deux points fondamentaux
- Le concept de littératie :
 - « La capacité des élèves à utiliser leurs connaissances dans des situations de la vie quotidienne, et à analyser, raisonner et communiquer de manière efficace. »
 - ✓ une dimension très large
 - ✓ qui ne repose pas sur les programmes scolaires
- La population visée :
 - ✓ 15 ans : fin de scolarité obligatoire
 - ✓ Une génération et non un niveau scolaire

Contenu

- Trois domaines évalués
- Cycles :

	2000	2003	2006	2009	2012	2015
Compréhension de l'écrit	<u>majeure</u>	mineure	mineure	<u>majeure</u>	mineure	mineure
Culture mathématique	mineure	<u>majeure</u>	mineure	mineure	<u>majeure</u>	mineure
Culture scientifique	mineure	mineure	<u>majeure</u>	mineure	mineure	<u>majeure</u>

- Chaque domaine comporte des sous-domaines
- Comparabilité dans le temps
 - ✓ Elle est mieux assurée entre deux dates en mode majeur

Exemple d'item

- Culture mathématique



L'image montre les traces de pas d'un homme en train de marcher. La longueur de pas L est la distance entre l'arrière de deux traces de pas consécutives. Pour les hommes, la formule $\frac{n}{L} = 140$ donne un rapport approximatif entre n et L , où : n = nombre de pas par minute, L = longueur de pas en mètres.

Question A :

Si la formule s'applique à la façon de marcher d'Henri et qu'Henri fait 70 pas par minute, quelle est la longueur de pas d'Henri ? Montrez vos calculs.

Quelques taux de réussite à cet item en 2006 :

OCDE (37%), Brésil (14%), France (43%), Hong-Kong (62%),
Russie (54%), Etats-Unis (28%)

Les élèves de 15 ans

- PISA 2009 :
 - ✓ les élèves nés en 1993
 - ✓ à des niveaux de scolarité différents, en raison des redoublements

	Classe fréquentée	Répartition (en %)
En avance	1 ^{ère} générale et technologique	2.5
"À l'heure"	2 ^{de} générale et technologique	51.4
	2 ^{de} professionnelle	9.2
En retard	3 ^e	31.9
	4 ^e	3.6
	Autre ou inconnu	1.4

Un mot sur la modélisation

- Modélisations statistiques sophistiquées : domaine de la psychométrie
- Modèle de mesure
 - ✓ Hypothèse 1 : L'objet mesuré est unidimensionnel : chaque élève peut être positionné sur un continuum en fonction de ses réponses au test
 - ✓ Hypothèse 2 : Cette dimension est invariante selon les pays, elle ne dépend pas des caractéristiques culturelles, linguistiques, sociales des élèves
- Le test est élaboré de manière à satisfaire au mieux ces hypothèses : la mesure est un « construit »
- Cependant, elles ne sont pas respectées : des « biais » persistent et sont intéressants à analyser

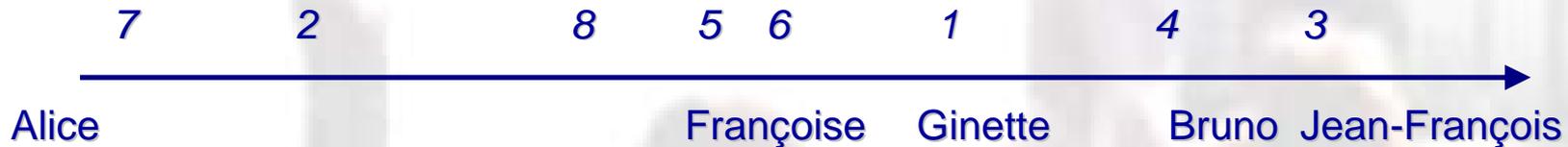
Modèle de mesure (1)

- Mesure d'une variable *latente*
- Ex : Mesurer la taille des individus...
...avec un questionnaire

- 1 Je dois souvent faire attention à ne pas me cogner la tête
 - 2 Pour les photos de groupe, on me demande souvent d'être au premier rang
 - 3 On me demande souvent si je fais du basket-ball
 - 4 Dans la plupart des voitures, je suis mal assis(e)
 - 5 Je dois souvent faire faire les ourlets quand j'achète un pantalon
 - 6 Je dois souvent me baisser pour faire la bise
 - 7 Au supermarché, je dois souvent demander de l'aide pour attraper des produits en haut des gondoles
 - 8 A deux sous un parapluie, c'est souvent moi qui le tiens
- ...

Modèle de mesure (2)

- Le score obtenu est fortement corrélé avec la taille réelle (*validité*)
- A quelques exceptions près, les items forment un ensemble homogène (*fidélité*)
- Construction d'une *échelle*



- L'échelle est unidimensionnelle
- L'échelle n'a pas d'unité ni de 0 absolu (comme les échelles de température)
- PISA transforme l'échelle : moyenne 500 et écart-type 100 pour l'ensemble des pays de l'OCDE

A blurred background image of a classroom. A teacher in a white shirt is standing at the front, and several students are seated at desks, some looking towards the front. The image is intentionally out of focus to emphasize the text overlay.

PISA

Organisation et méthodes

“Comitologie”

- Structure :

- ✓ Le pilote : Secrétariat de l'OCDE
- ✓ Les décisionnaires : PISA Governing Board (règle du « consensus » entre pays autour des propositions du Secrétariat)
- ✓ Le chef d'orchestre : le consortium international dirigé par ACER (Australian Council for Educational Research) depuis 2000 (changement pour 2015)
- ✓ Les groupes d'experts (REG, MEG, SEG, QEG, TAG)
- ✓ Les exécutants : les « centres nationaux » (services statistiques, universités ou prestataires privés)

Coûts

- Coûts :
 - ✓ Participation aux coûts internationaux (au total, 10M €)
 - ✓ Coûts nationaux équivalents
 - ✓ Enquêtes cycliques : mouvement perpétuel
 - ✓ Possible collusion avec les programmes nationaux d'évaluation

Contraintes

- « Cahier des charges » :
 - ✓ Liste de 56 « standards » (taux de réponse, passation, correction, traduction, gestion des données, etc.)
 - ✓ Contrôle-qualité à tous les niveaux (centres nationaux, établissements) : détection des erreurs, voire des fraudes
 - ✓ Punitons pour non-respect des standards : du « flag » à l'exclusion
- Un standard important, le taux de réponse :
 - ✓ 85% des établissements visés (min. 150), 80% des élèves parmi les établissements répondants (min . 4500)
 - ✓ Exclusion des Pays-Bas en 2000 et du Royaume-Uni en 2003 pour le non-respect de ce standard
 - ✓ En France, excellent taux de réponse des établissements

Population

- Nuance importante dans certains pays :
 - ✓ PISA : élèves de 15 ans, scolarisés au moins au grade 7 (5^{ème} en France)
 - ✓ i.e. sont écartés d'emblée : 5% en Corée du Sud, 37% au Mexique, 13% au Portugal, 43% en Turquie, 13% en Argentine, 30% au Brésil, 29% en Roumanie, 19% en Thaïlande
- Des exclusions sont autorisées :
 - ✓ Exclusions d'établissements (inaccessibles, spécialisés)
 - ✓ Exclusions d'élèves (déficiences intellectuelles, fonctionnelles, primo-arrivants)
 - ✓ Standard : exclusions < 5%
 - ✓ En France : exclusion des TOM, de la Réunion, des EREA, et des établissements privés hors contrats

Echantillonnage

- Sondage :
 - ✓ Tirage au sort d'un échantillon d'établissements (en France, environ 180 selon les vagues d'enquêtes)
 - ✓ Tirage de 32 élèves de 15 ans au plus dans chacun des établissements sélectionnés (N=4500 au total en France)
- Erreurs et intervalles de confiance:
 - ✓ Nécessité de tenir compte des erreurs d'échantillonnage pour conclure à la significativité des résultats
 - ✓ Exemple : pour conclure à une évolution significative du score moyen de la France entre deux années, il faut observer une différence d'au moins 13 points entre les deux scores
 - ✓ Très important à retenir pour l'interprétation des tendances ou des rangs

Elaboration des tests

- Développement des items
 - ✓ Année n-2 : contributions du consortium et des pays
 - ✓ Année n-1 : expérimentation des items sur des échantillons réduits (environ 1200 élèves par pays)
 - ✓ Année n : passation des items sélectionnés pour l'étude principale
 - ✓ PISA 2006 sciences : 492 items créés, 247 conservés pour l'expérimentation de 2005 (dont 37% proposés par les pays), 92 retenus pour l'enquête principale de 2006
 - ✓ Au total, l'équivalent de 6h30 de matériel d'évaluation est réparti dans des cahiers « tournants » (2h chacun)

Processus empirique (un « construit ») et « procrustéen »

Traduction

- Procédures :
 - ✓ 44 langues, 87 versions (10 versions en anglais, 5 en français)
 - ✓ Principe : assurer un niveau de difficulté équivalent dans toutes les langues
 - ✓ Tous les items sont traduits par la consortium en deux langues « source » : l'anglais et le français
 - ✓ Les pays doivent suivre des procédures de double traduction avec conciliation par un tiers
 - ✓ Au final : des fonctionnements différentiels subsistent selon la langue (si un item A est plus difficile qu'un item B en France, cela reste vrai le plus souvent dans les pays francophones, mais pas forcément dans les autres pays)

Administration de l'enquête

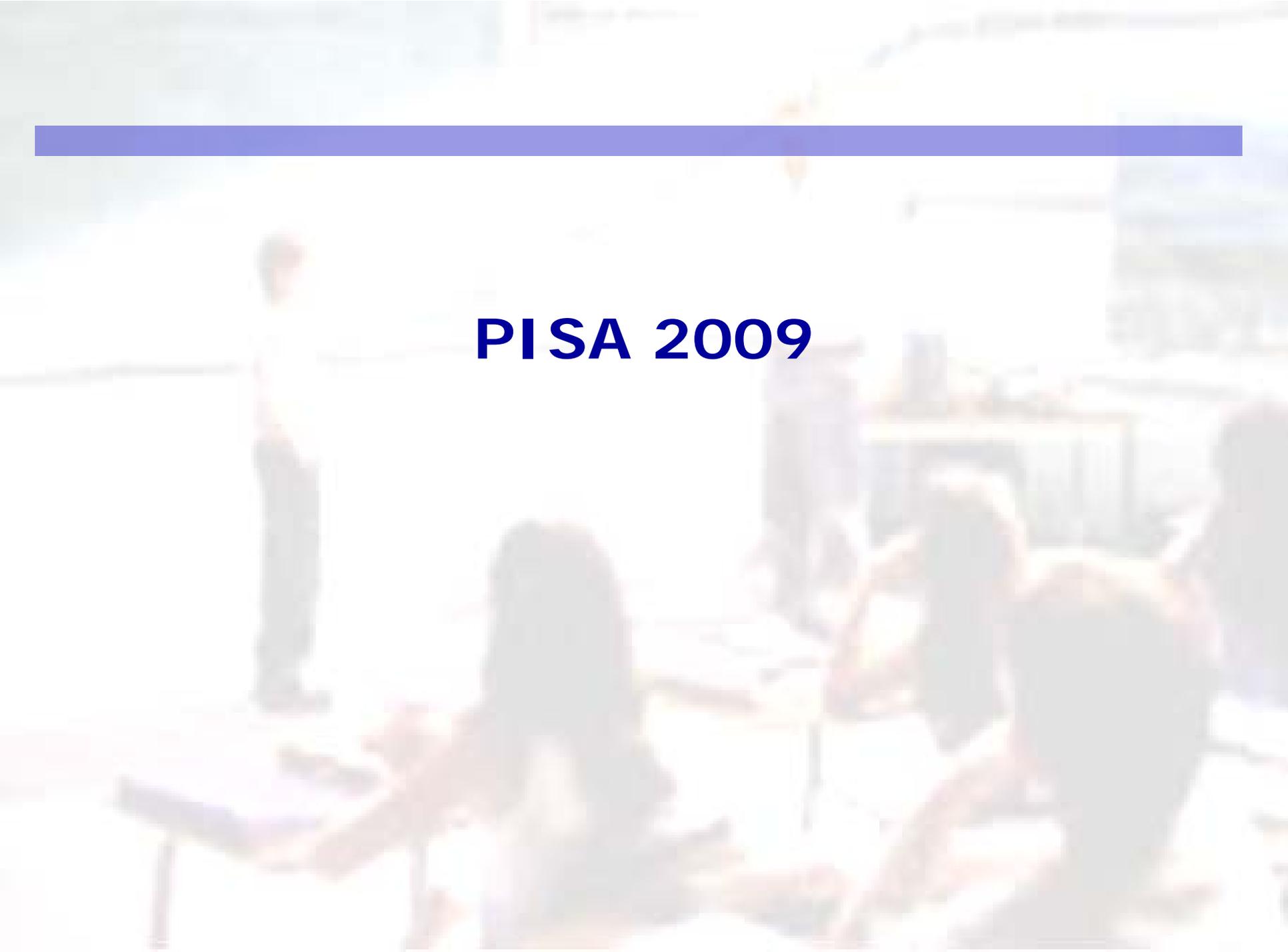
- Passation :
 - ✓ 2h d'épreuves + 30' de questionnaire « contextuel »
 - ✓ Administrateurs de test externes : pas en France (en général, l'adjoint et/ou un CPE)
 - ✓ En 2015 : passation par ordinateur
- Maximiser la participation :
 - ✓ Etablissements de remplacement, sessions de rattrapage
 - ✓ Certains pays paient les établissements et/ou les élèves
 - ✓ Cas difficiles : grève, boycott
- Une question importante :
 - ✓ l'implication et la motivation des élèves dans une évaluation « sans enjeux » pour eux
 - ✓ procédures pour favoriser la motivation : expérimentation en 2011

Correction des réponses

- Procédures :

- ✓ PISA 2009 : 34% de questions ouvertes, appelant une réponse rédigée de la part de l'élève
- ✓ Consignes de correction précises (soumises également au processus de traduction)
- ✓ Formation des correcteurs (pas forcément des enseignants)
- ✓ Quadruple corrections sur un sous-échantillon de 100 élèves, par questions
- ✓ Taux de cohérence élevé (concentration des erreurs sur des questions, plutôt que sur des correcteurs)

Un mot sur la docimologie

A blurred photograph of a classroom. A teacher in a white shirt stands at the front, and several students are seated at desks, some looking towards the front. A thick purple horizontal bar is at the top of the image.

PISA 2009

Résultats PISA 2009

Les Echos
LE QUOTIDIEN DE L'ÉCONOMIE

ÉDUCATION

Dans le dernier test Pisa, qui évalue les élèves de quinze ans, la France, tout en restant dans la moyenne, chute de la 12^e à la 18^e place en compréhension de l'écrit et du 13^e au 16^e rang en mathématiques. En cause, un système inégalitaire qui engendre toujours plus d'échec scolaire.

Mauvaise note pour la France, élève moyenne mais championne de l'échec scolaire



ÉDUCATION L'étude Pisa 2009 de l'OCDE souligne de fortes disparités de niveaux entre les élèves

LA FRANCE, UNE ÉLÈVE QUI NE PROGRESSE PAS

Le Monde

L'école française tout juste dans la moyenne des pays de l'OCDE

L'évaluation du niveau des jeunes Français montre que près d'un quart des élèves est à la traîne

L'EXPRESS

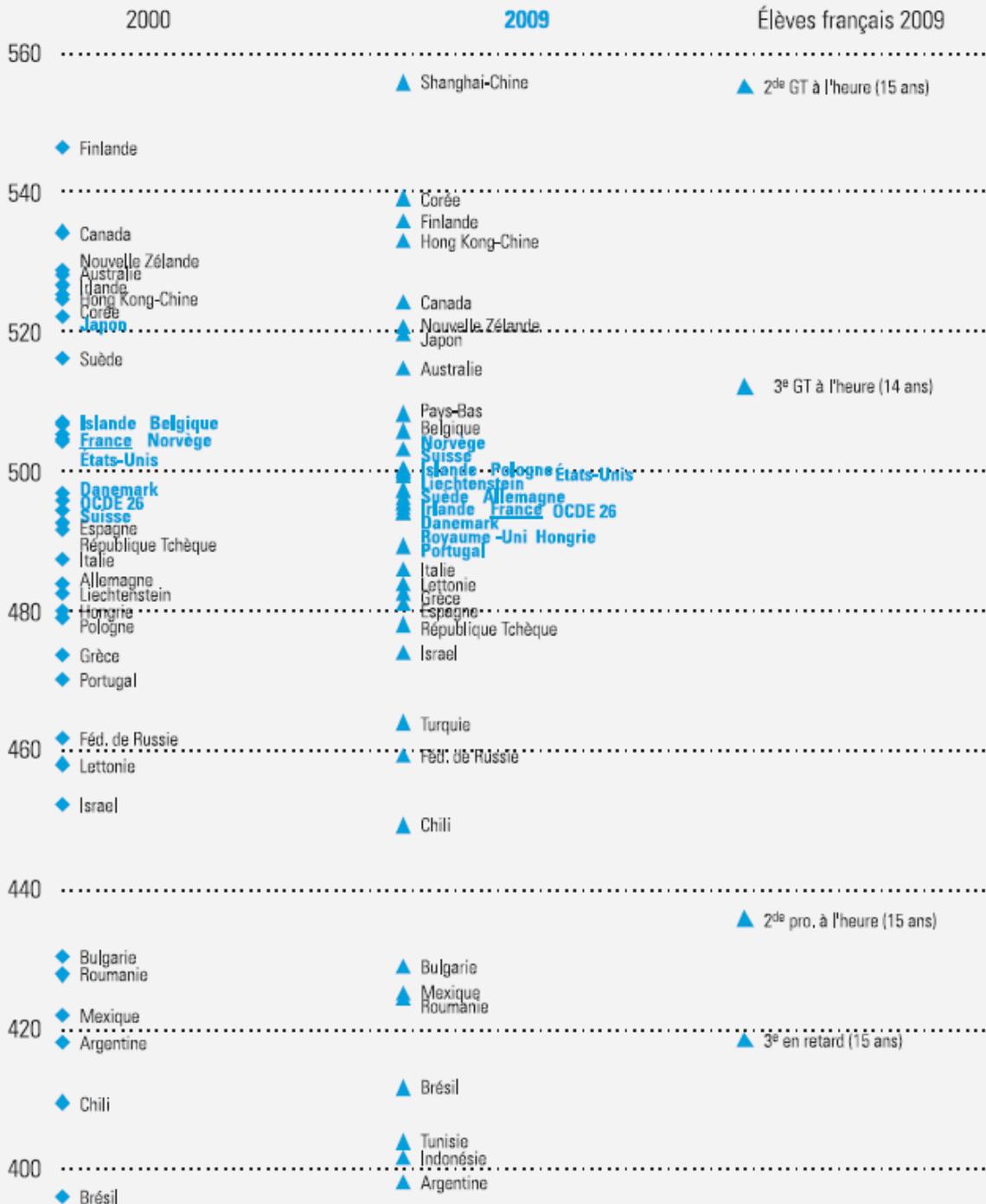
La France, élève moyenne

L'OCDE a noté notre système scolaire parmi ceux de 65 pays. Verdict : « Doit mieux faire », surtout pour les plus en difficulté. Commentaire de copie par un expert.

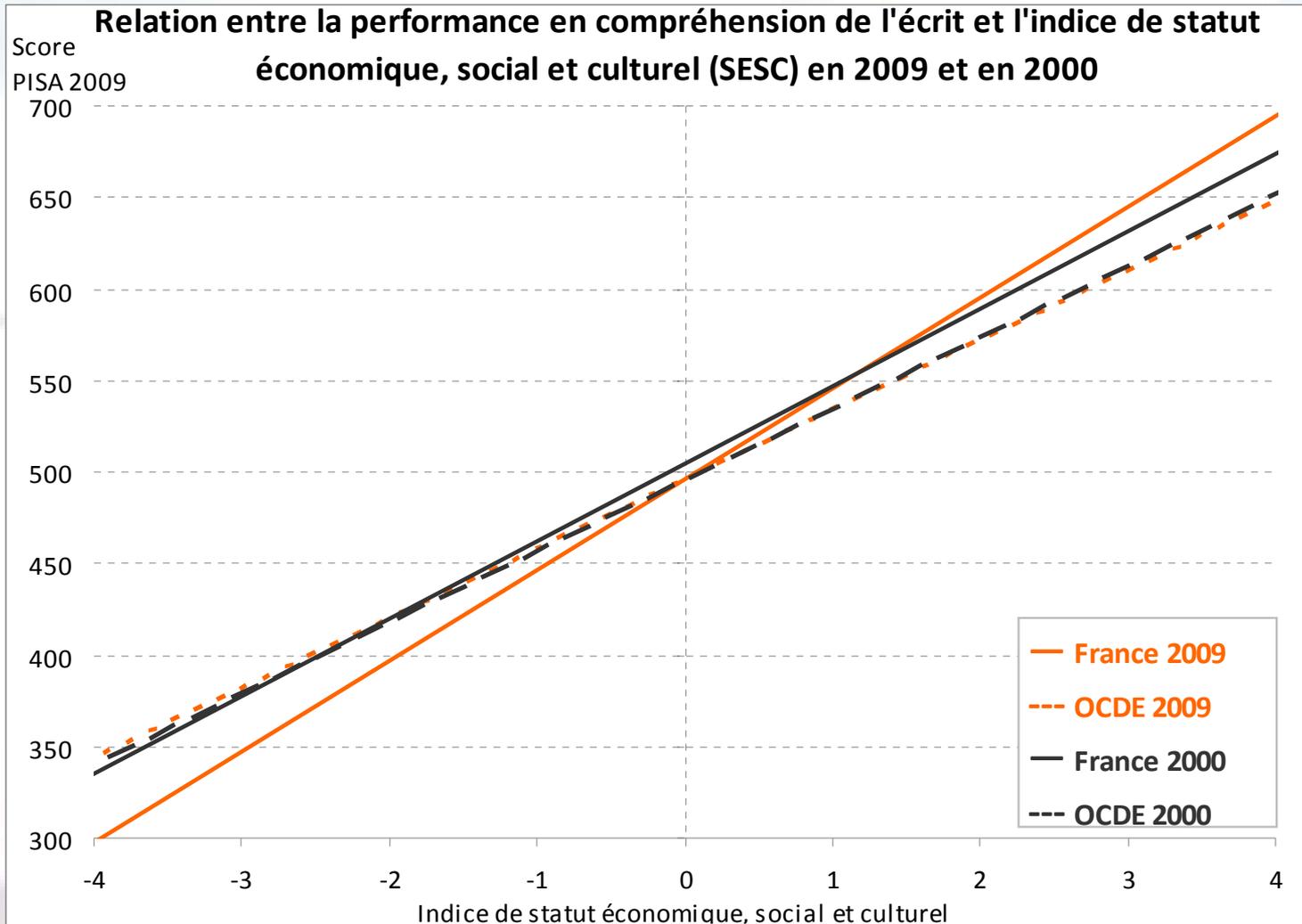
Quelques remarques pour l'analyse

- Le classement :
 - ✓ Le rang est un indicateur fragile : en tenant compte des erreurs, la France est entre la 11^e et la 22^e place sur les 34 pays de l'OCDE dans PISA 2009
 - ✓ Les pays : ensemble de référence (OCDE, tous), Shanghai=Chine ?
- La significativité
 - ✓ Il faut tenir compte de la significativité statistique (indiquée dans les rapports)
- Les distributions
 - ✓ Au-delà du score moyen, des niveaux de performance ont été établis et permettent d'affiner le diagnostic
- Lien avec des variables socio-démographiques
 - ✓ Les variables « contextuelles » sont des indices composites, basés sur les déclarations des élèves

PISA 2009 - 1

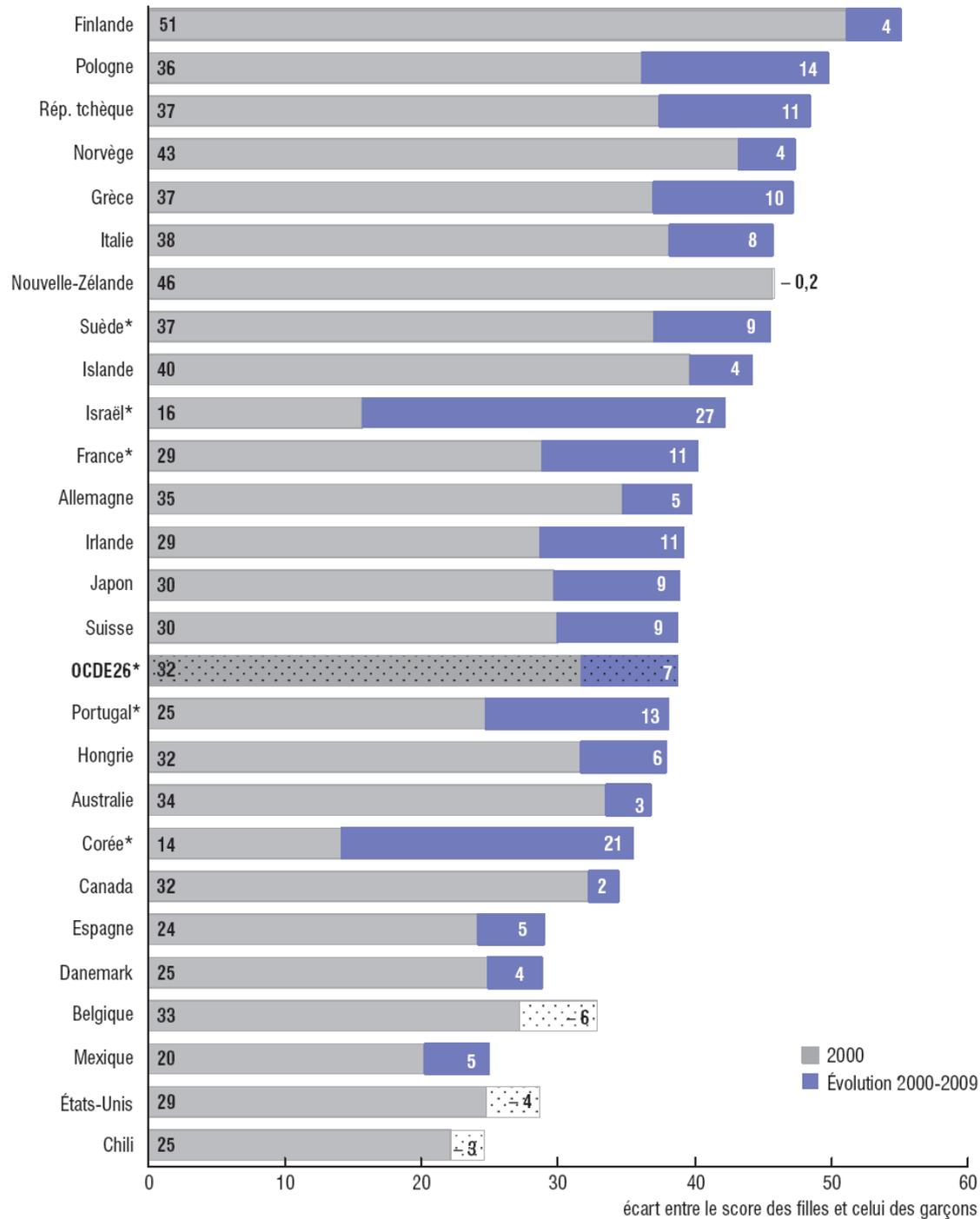


PISA 2009 - 2



PISA 2009 - 3

- Garçons / filles



► Évaluer, mais pas trop...

On ne pourra que regretter qu'une étude aussi précise que Pisa ne puisse préciser certains détails concernant l'école française. Mais la France est le pays qui fournit le moins de données à l'OCDE. Curieusement, les performances par établissement d'origine ne sont pas renseignées, la France étant la seule à exercer son droit de retrait sur ce sujet. Impossible de connaître le rapport entre les performances en lecture des élèves et le statut, favorisé ou défavorisé, des établissements. Selon le ministère, les « différences entre établissements » pour les élèves de 15 ans rendraient les données non significatives. ■

Enquête sur le niveau scolaire: les cachotteries de la France

La France aime se distinguer. Or dans la dernière enquête Pisa de l'OCDE, les résultats de ses élèves sont désespérément moyens. Mais elle a quand même réussi à se faire remarquer: parmi les 65 pays ayant participé à l'étude, elle est en effet la seule à avoir refusé de répondre à une série de questions, celles sur les établissements scolaires. Ce qui expli-

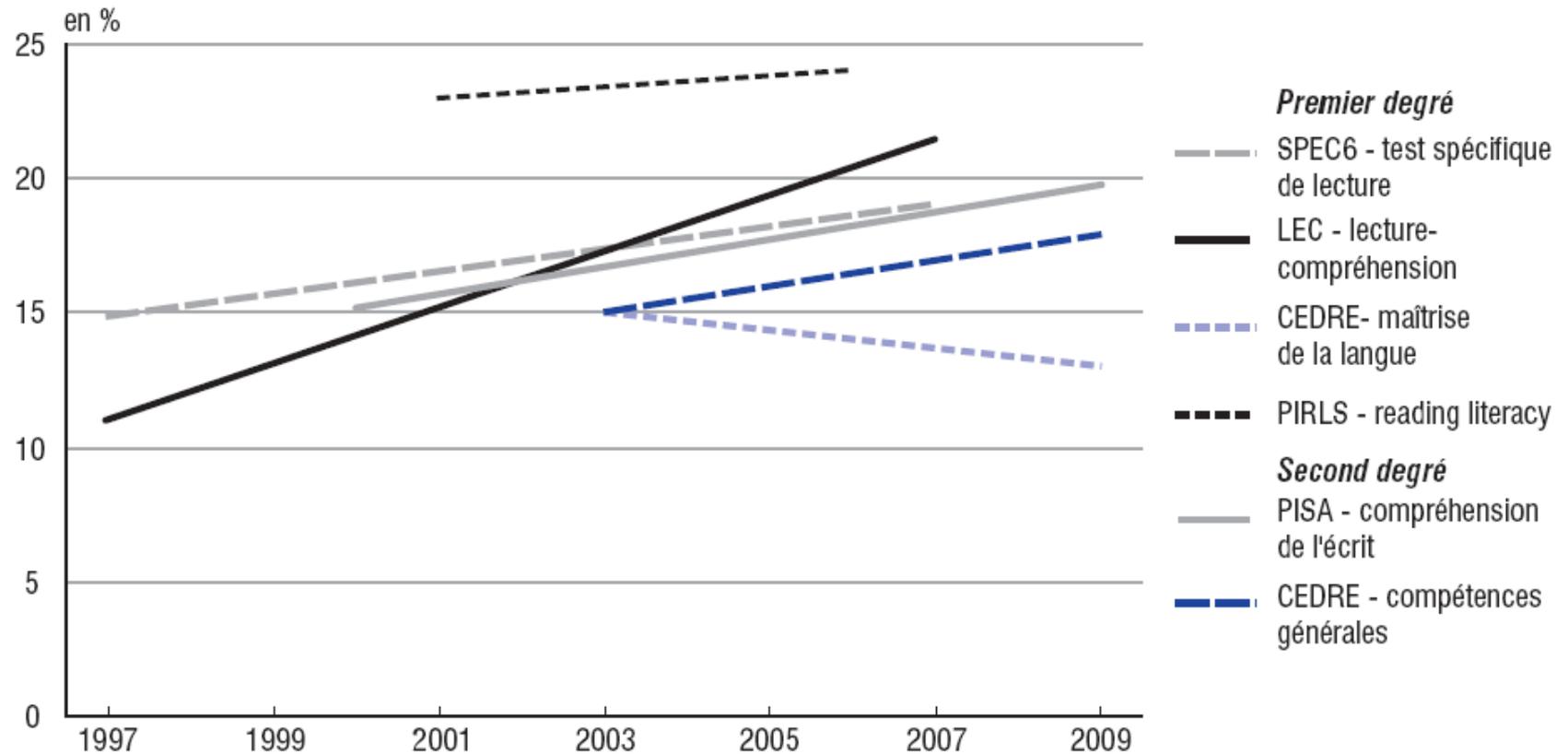
viseurs à y répondre? On aurait probablement mesuré ce que l'on sait déjà. Dans certains établissements, l'absentéisme est important, ainsi que les incivilités à l'encontre des profs ou le harcèlement entre élèves. Les profs n'ont pas très bon moral et n'attendent pas grand-chose des élèves. Et lorsqu'ils sont malades, ils tardent à être remplacés. A

• Le problème :

- ✓ En France, du fait du redoublement et de la coupure collège/lycée à 15 ans, les élèves de l'échantillon ne sont pas représentatifs des élèves de leurs établissements
- ✓ Comparer les établissements dans PISA revient à comparer les élèves en retard de collège avec les élèves à l'heure de lycée
- ✓ Par conséquent, les différences inter-établissements sont sur-estimés dans PISA (50% de variance expliquée, contre environ 20-25% en réalité)

En complément (1)

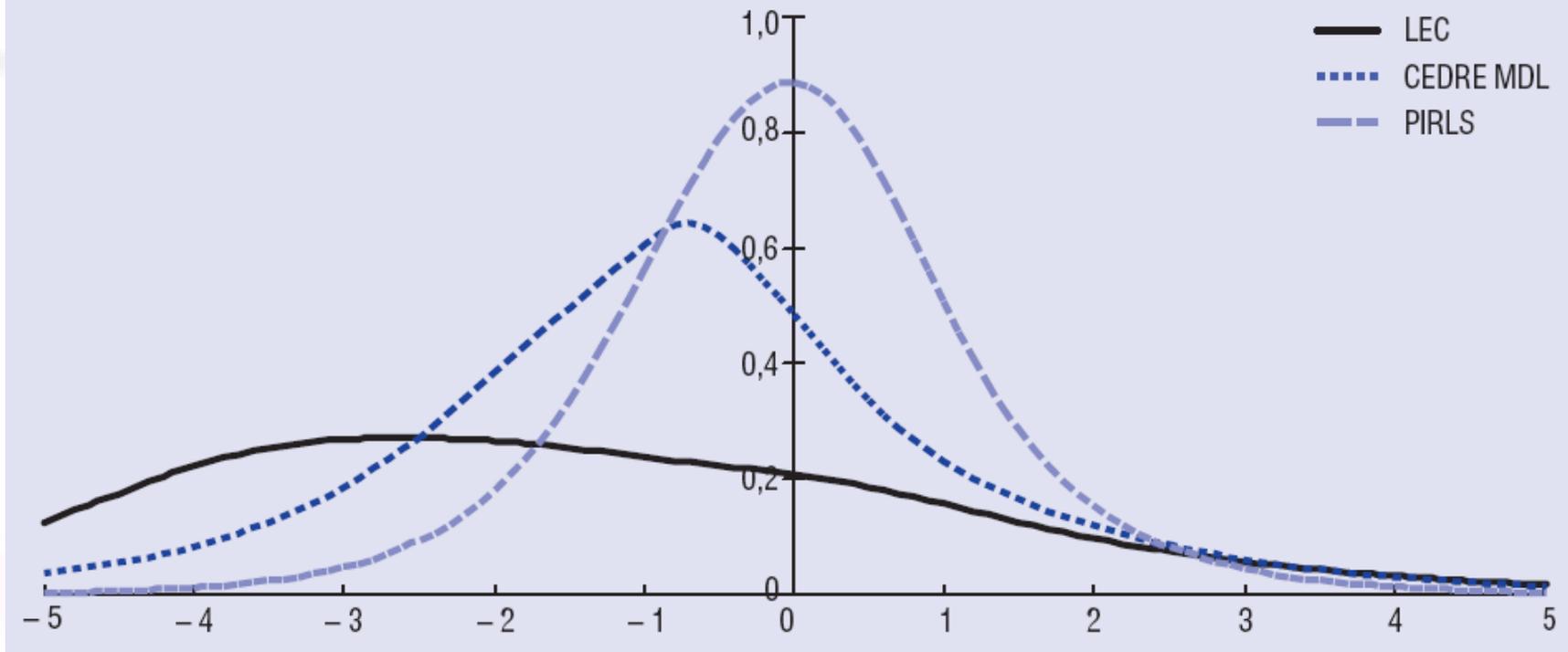
- Elèves en difficulté face à l'écrit



En complément (2)

- A chaque test sa précision selon le niveau de compétence

Information des tests



Ressources

- Résultats PISA 2009:
 - ✓ Le site de l'OCDE : www.pisa.oecd.org
 - ✓ Les notes de la DEPP sur www.education.gouv.fr
dans Outils de documentation et d'information / Les publications / Notes d'information
- Actualité :
 - ✓ Résultats de PIRLS 2011
 - ✓ Conférence de presse le 11 décembre 2012

ANNEXE 2

Les enquêtes internationales en France : de l'indifférence à la « référence ».

Au risque d'être quelque peu simplificateur, c'est ainsi que l'on peut caractériser l'évolution des réactions françaises au cours de la dizaine d'années écoulée depuis la première publication des résultats de PISA : les responsables, les acteurs et les partenaires de la politique éducative française, après avoir estimé et soutenu que cette enquête était peu pertinente pour rendre compte des résultats de notre système éducatif, en font aujourd'hui une référence obligée, voire une caution, pour justifier leurs politiques, leurs positions ou leurs revendications, mais ils ne prennent pas vraiment en compte les questions que les résultats de PISA suscitent quant à notre système éducatif, pour piloter celui-ci.

Cette appréciation ne concerne pas que PISA ; elle rejoint celle, plus générale, que l'on peut faire concernant « *la transposition, pour le moins limitée, des résultats ... élaborés dans l'univers de la recherche, dans les univers de la décision politique et de l'action pédagogique* ». Pour la préciser, il convient de « *se pencher sur les deux parties prenantes de cette interaction* », d'une part « le « *producteur de la connaissance* », en l'occurrence l'OCDE, organisation intergouvernementale et les experts auxquels elle fait appel pour réaliser et exploiter l'enquête PISA et, d'autre part, les « *récepteurs potentiels de ces savoirs* »¹, en l'occurrence les responsables et les acteurs de la politique éducative, et plus largement le corps social.

S'agissant de PISA, le statut du « *producteur* » qu'est l'OCDE, donne à cette enquête un caractère et un retentissement particuliers au sein de l'ensemble des connaissances issues des travaux de recherche et d'évaluation du système éducatif. La première enquête PISA a eu lieu en 2000, alors que l'évaluation des systèmes d'enseignement par le recours à de grandes enquêtes de masse sur les compétences et les acquis des élèves s'inscrit dans une « *histoire de cinquante ans* »². Mais jusqu'à ce que l'OCDE s'en empare, ces enquêtes étaient réalisées par des associations scientifiques internationales³ et ne rencontraient généralement que très peu d'écho auprès des responsables politiques et du grand public.

C'est la prise de conscience des difficultés – voire de la crise – que connaissaient les systèmes d'enseignement, notamment aux Etats-Unis, qui a conduit l'OCDE à recueillir et publier régulièrement un ensemble cohérent d'indicateurs internationaux de l'enseignement. La décision de réaliser PISA s'est inscrite dans la continuité de cette initiative : il est rapidement apparu indispensable, tant aux responsables politiques qu'aux « *statisticiens d'État* », de structurer cet ensemble d'indicateurs autour des résultats⁴.

¹ Agnès Van Zanten, *Le rôle de la connaissance dans la régulation du système éducatif en France : questions de production, questions de réception*, in Actes du colloque « le pilotage du système éducatif, enjeux, outils, perspectives » octobre 2004, MEN/DEP, PARIS.

² Norberto Bottani, Pierre Vrignaud, *La France et les évaluations internationales*, rapport n°16, établi à la demande du Haut Conseil de l'évaluation de l'école, janvier 2005, PARIS.

³ Il s'agit tout particulièrement de l'IEA (International Association for the Evaluation of Educational Achievement), créée en 1961, et qui a réalisé de nombreuses enquêtes en développant les méthodologies qui sont utilisées aujourd'hui, notamment pour PISA.

⁴ L'histoire et les enjeux du développement de cet ensemble d'indicateurs (Regards sur l'éducation, publié chaque année par l'OCDE depuis 1991), sont retracés par Norberto Bottani dans *Nouveaux savoirs sur l'enseignement, enjeux sur le plan international*, in Évaluer l'évaluation, actes du colloque organisé par l'IREA, Éditions Le Manuscrit, 2009, PARIS.

Issue d'initiatives gouvernementales, financée par les gouvernements⁵, pilotée par une organisation intergouvernementale importante comme l'OCDE, dont c'est un des « produits phares » et qui mobilise toute sa puissance de communication pour en diffuser les résultats, PISA ne peut laisser indifférents les « récepteurs » que sont les responsables politiques et les opinions publiques. Et, depuis 2000, le Ministère de l'Éducation nationale, poussé par l'OCDE et sollicité par les médias, a toujours dû prendre position au regard des résultats de l'enquête qui introduisaient dans le débat public une nouveauté particulièrement attrayante : le classement des pays selon les acquis de leurs élèves en fin de scolarité obligatoire. Dans un premier temps, c'est ce seul classement qui a retenu l'attention. Les concepteurs de l'enquête ont d'ailleurs largement mis en avant cet aspect, très médiatique mais réducteur, voire trompeur. Ils affirmaient que « *la dimension internationale [du PISA] permet aux décideurs d'infirmer ou de confirmer des hypothèses concernant la qualité du rendement scolaire de leur pays* », tout en prenant quelques précautions formelles en rappelant que « *si les résultats obtenus par un pays donné dans le cadre du PISA sont supérieurs à ceux d'un autre pays, cela ne signifie pas pour autant que les établissements du premier sont plus performants. En effet, l'apprentissage commence bien avant la scolarisation et se poursuit dans divers environnements institutionnels et extrascolaires. En revanche, l'on peut en conclure avec raison que, dans ce pays, l'impact cumulé de toutes les expériences d'apprentissage vécues entre la petite enfance et l'âge de 15 ans, à l'école et en dehors de l'école, a généré des résultats plus favorables dans les domaines évalués par le PISA*⁶.

2000 : résultats honorables, discrétion et prise de distance.

Les réactions françaises à la première vague de PISA ont été relativement discrètes. Elles ont insisté sur le fait que le cadre théorique sur lequel était fondé PISA⁷ rendait mal compte des objectifs de notre enseignement. Elles l'ont fait d'autant plus facilement que la position française dans le « palmarès » international n'était pas catastrophique, même si notre système éducatif n'y apparaissait pas (ou plus ?) comme l'un des meilleurs du monde : « *En Europe, la France se situe toujours devant l'Allemagne, arrive en tête des pays latins et devance les pays d'Europe de l'Est. Elle a des résultats du même ordre que les Etats-Unis, sauf en culture mathématique où son score est nettement supérieur* »⁸. Contrairement à l'Allemagne, la France n'a pas connu de « *PISA schock* » !

Ces réactions s'appuyaient sur des travaux de la DEP, le « producteur » français de PISA. Celle-ci ne s'était pas contentée d'organiser la passation des épreuves de PISA en France, mais avait proposé de jouer un rôle actif dans la préparation de l'enquête, en prenant l'initiative d'organiser un consortium international pour répondre à l'appel d'offres que l'OCDE avait lancé, en 1998, pour sélectionner le maître d'œuvre international de PISA. Cette réponse n'a pas été retenue, mais a eu des effets : ainsi, la grille des compétences de lecture, élaborée pour répondre à l'appel d'offres par une experte de l'INRP, a été reprise dans le cadre conceptuel utilisé pour évaluer les compétences en lecture dans PISA.

⁵ PISA coûte cher, de l'ordre de 460 000 € par an, dont 270 000 € versés à l'OCDE et environ 190 000 €, qui s'imputent sur les moyens humains et financiers de la DEPP.

⁶ Ces citations proviennent du rapport sur les premiers résultats de PISA : *Connaissances et compétences, des atouts pour la vie. Premiers résultats de PISA 2000*, OCDE 2001, PARIS. C'est moi qui souligne.

⁷ *Le PISA se concentre sur ce dont les jeunes de quinze ans auront besoin dans leur vie future et son propos est d'évaluer ce qu'ils vont pouvoir faire grâce à ce qu'ils ont appris. Le PISA propose une nouvelle manière d'envisager le rendement scolaire sur la base des expériences des élèves dans le monde, plutôt que dans le contexte culturel spécifique d'un seul pays.* (Ibid.)

⁸ « *Les élèves de 15 ans – Premiers résultats d'une évaluation internationale des acquis des élèves*, Note d'information 01.52, décembre 2001, MEN-DPD.

Après la publication des premiers résultats, la DEP a, à la fois, développé un ensemble de remarques critiques sur la méthodologie de PISA : prétention à développer un modèle universel de compétences, existence de biais culturels, etc.,⁹ financé quelques recherches sur les outils et les résultats de PISA et diffusé de premières analyses qui interrogeaient nos objectifs éducatifs et nos pratiques¹⁰. Quelques chercheurs, en nombre limité, ont commencé à étayer leurs travaux en utilisant les résultats de PISA, alors que l'Inspection générale de l'Éducation nationale ne s'y intéressait pas particulièrement.

2003 : discrétion toujours, mais premières réflexions.

Les résultats de la vague suivante de PISA, en 2003, consacrée plus particulièrement à la culture mathématique, n'ont guère eu plus d'écho public que ceux de la première. Ceci, d'autant plus qu'« *au sein des trente pays de l'OCDE, la France obtient un score global de Culture mathématique significativement au-dessus de la moyenne de ces pays* » et qu'« *en Compréhension de l'écrit et en Culture mathématique, le score des élèves français se maintient par rapport aux résultats d'il y a trois ans, tandis qu'en Culture scientifique il progresse légèrement* »¹¹. On constatait ainsi, une fois de plus, que lorsque les résultats d'une évaluation ne sont pas (trop ?) mauvais, il est inutile d'en parler plus avant, sauf à s'appuyer sur ces résultats pour justifier la continuité ou l'inflexion d'une politique.

Ce sont d'ailleurs moins les résultats de la France relativement aux autres pays, que la conception des acquis éducatifs qui sous-tendait PISA, qui a contribué à nourrir les réflexions qui ont conduit à définir « *un socle commun constitué d'un ensemble de connaissances et de compétences qu'il est indispensable de maîtriser pour accomplir avec succès sa scolarité, poursuivre sa formation, construire son avenir personnel et professionnel et réussir sa vie en société* »¹². Ces réflexions s'inscrivaient dans un mouvement international consistant à demander aux systèmes éducatifs de doter les élèves de « compétences-clefs » en fin de scolarité obligatoire. Toujours est-il que le décret qui présente la conception et le contenu du socle commun précise que la définition du socle « *se réfère ... aux évaluations internationales, notamment au PISA qui propose une mesure comparée des connaissances et des compétences nécessaires tout au long de la vie.* »¹³. C'est certainement la première fois qu'un texte définissant des objectifs de notre École fait une référence explicite à une évaluation standardisée, au surplus internationale, des acquis des élèves.

Les données de PISA 2003 ont également permis de constater qu'« *alors que globalement, la France se situe au-dessus de la moyenne internationale en culture mathématique, les élèves français « à l'heure » à 15 ans et scolarisés en seconde générale et technologique obtiennent un score supérieur [...] aux scores moyens des élèves de 15 ans de Hong Kong ou de Finlande, pays qui arrivent en tête du palmarès international. À l'inverse, les élèves ayant un an de retard et scolarisés à 15 ans en troisième générale ont un score [très] inférieur [qui les] situe au niveau des scores moyens du Portugal et de l'Italie [...]. On retrouve des différences de même ampleur sur les échelles de Compréhension de l'écrit,*

⁹ Ces critiques ont largement été reprises par le Haut Conseil de l'évaluation de l'école dans son avis n°16, *la France et les évaluations internationales des acquis des élèves*, mai 2005 : Il les résumait ainsi : *Sans entrer ici dans le détail, il faut être conscient que la mesure des compétences des élèves (et des adultes) n'a pas la simplicité, la robustesse et l'universalité du système métrique. Dans le domaine de la psychométrie – ou pour mieux dire de l'éduométrie, c'est à dire de la mesure en éducation – les mesures sont des constructions qui résultent de choix politiques et méthodologiques. Ces choix peuvent aussi avoir des dimensions commerciales.*

¹⁰ Voir en particulier : *Les compétences des élèves français à l'épreuve d'une évaluation internationale, Premiers résultats de l'enquête PISA 2000*, Les Dossiers évaluations et statistiques, n° 137, novembre 2002, MEN-DPD, PARIS.

¹¹ *Les élèves de 15 ans – Premiers résultats de l'évaluation internationale PISA 2003*, Note d'évaluation n° 04-12, décembre 2004, MENESR/DEP, PARIS.

¹² Article L 122-1-1 du code de l'éducation.

¹³ Décret n° 2006-830 du 11 juillet 2006 relatif au socle commun de connaissances et de compétences.

de Sciences et de Résolution de problèmes¹⁴ ». Ce constat qui appelait de sérieuses réflexions sur la pratique française du redoublement¹⁵ n'a pas été pris en compte à l'époque¹⁶, il l'a été plus tard, sans doute autant ... sous la contrainte budgétaire, que par conviction.

2006, puis 2009 : comment gérer des résultats en baisse ?

La publication des résultats de PISA 2006 marque une inflexion dans la réception de PISA. C'est la troisième vague de l'enquête, et elle permet des comparaisons temporelles des acquis des élèves. Celles-ci ne sont pas flatteuses pour notre système éducatif : « depuis 2000, les résultats de la France indiquent une tendance à la baisse, plus par rapport à elle-même que par rapport à la moyenne des pays de l'OCDE. En effet, la moyenne des pays de l'OCDE baisse légèrement depuis 2000, alors que la baisse de la France est plus accentuée. Les élèves des bas niveaux sont plus nombreux et leur performance est plus faible. [...] Les difficultés de lecture et d'écriture pèsent sur la réussite globale à PISA¹⁷ ».

Ce constat sera confirmé par les résultats de PISA 2009 qui permettent, pour la première fois, de juger de l'évolution des performances en compréhension de l'écrit sur une période longue, de 2000 à 2009. Il est convergent avec les analyses issues des évaluations nationales réalisées par la DEPP¹⁸.

Les responsables politiques ne peuvent plus ignorer ces résultats, parce leur publication est devenue aujourd'hui un événement médiatique et politique incontournable, étant donné l'audience qu'a su lui gagner l'OCDE. Cette médiatisation a d'ailleurs profité aux autres enquêtes internationales, notamment celles régulièrement réalisées depuis plusieurs dizaines d'années par l'IEA (PIRLS, TIMSS, ...) qui, avant 2000, n'étaient guère connues que des spécialistes qui les avaient conçues.

Ils ne peuvent pas non plus les écarter parce qu'« on s'accorde, au sein même de la communauté des chercheurs en éducation, sur le fait que le programme de l'OCDE a bénéficié d'une expertise et d'une attention méthodologique qui obligent à déplacer la critique sur un autre terrain que le caractère plus ou moins rigoureux de l'enquête elle-même », et ceci, même s'il faut « souligner les limites de ce type d'enquêtes à qui on veut parfois faire dire plus qu'elles ne peuvent signifier »¹⁹.

Ainsi, en France, les comparaisons internationales sont de plus en plus évoquées par les responsables politiques, mais elles permettent plutôt de justifier une décision, qu'elles ne l'éclairent. Les résultats de l'enquête PIRLS (enquête sur la lecture en fin de primaire) qui donnent une image moyenne de notre enseignement primaire, ont ainsi été évoqués lors de la décision de réduire à 24 heures la semaine des écoliers, alors que l'analyse des résultats – peut-être insuffisante - ne disait rien sur l'influence, et encore moins sur le sens de l'influence, que pouvait avoir le temps d'enseignement sur les résultats.

¹⁴ Les élèves de 15 ans – Premiers résultats de l'évaluation internationale PISA 2003, Note d'évaluation n° 04-12, décembre 2004, MENESR/DEP, PARIS.

¹⁵ Voir notamment Olivier COSNEFROY et Thierry ROCHER, *Le redoublement au cours de la scolarité obligatoire : nouvelles analyses, mêmes constats*, dossier Education & formations, n°166, mai 2005, MENESR-DEPP, PARIS.

¹⁶ La Loi d'orientation pour l'avenir de l'École d'avril 2005 est plutôt favorable au redoublement.

¹⁷ *L'évolution des acquis des élèves de 15 ans en culture mathématique et en compréhension de l'écrit – premiers résultats de l'évaluation internationale PISA 2006*, Note d'information n°08-08, janvier 2008, MEN/DEPP, PARIS.

¹⁸ Jeanne-Marie Daussin, Saskia Keskaik, Thierry Rocher (DEPP) *L'évolution du nombre d'élèves en difficulté face à l'écrit depuis une dizaine d'années*. France, portrait social – INSEE références – Edition 2011, Paris.

¹⁹ *PISA : ce que l'on en sait et ce que l'on en fait*, par Olivier Rey, dossier d'actualité veille et analyse, n°66, octobre 2011, IFÉ, ENS de Lyon.

En décembre 2010, c'est le Ministre de l'Éducation nationale lui-même qui a présenté à la presse les résultats de PISA 2009, alors que ses prédécesseurs laissaient plutôt cette tâche aux fonctionnaires de la Direction de l'évaluation et de la prospective. Mais, après avoir présenté – de façon tout-à-fait objective – l'image peu flatteuse du système éducatif français que donne l'enquête, il s'est contenté de rappeler les grandes lignes de sa politique, à peu près de la même façon qu'il aurait pu le faire la veille.

Les enquêtes internationales, un matériau pour la réflexion et pour l'action.

Pourtant, si l'on prend un peu de distance par rapport aux seuls discours, on a vu que PISA, comme d'autres comparaisons internationales, a contribué à nourrir le débat éducatif français : définition d'un socle commun de compétences et de connaissances en fin de scolarité obligatoire, volonté politique plus affirmée aujourd'hui de limiter les redoublements, interrogations sur certains objectifs de notre enseignement de la lecture, sur le statut de l'erreur dans nos pratiques, sur la faible estime de soi, voire le *stress* de nos élèves, etc.. Mais nourrir le débat est une chose, encore faut-il que celui-ci débouche sur des orientations explicitées, effectivement partagées et mises en place.

Tient-on vraiment compte de ce que pourrait être l'apport de PISA ? Se pose-t-on la question : à moyens égaux d'autres font mieux, pourquoi ?

Citons simplement ces conclusions qui, à partir des résultats de PISA 2009, s'efforcent de répondre à la question : « *Qu'est-ce qui rend une École performante ?* »²⁰ :

« Les caractéristiques que l'on rencontre plus fréquemment dans les systèmes éducatifs dont les résultats sont les meilleurs, c'est à dire ceux qui associent efficacité (niveau élevé de résultats globaux) et équité (faible impact de la situation sociale des élèves sur leurs résultats) seraient les suivantes :

- peu de « différenciation », c'est-à-dire, pas d'orientation précoce dans des filières différentes selon les résultats des élèves ; des écoles où règne la mixité sociale ; peu, voire pas de redoublement ; peu d'exclusions ou de transferts entre écoles d'élèves « agités » ou aux faibles résultats ;*
- un degré d'autonomie des écoles plutôt important en matière de curriculum, associé à une faible compétition entre écoles quant au recrutement des élèves (autrement dit la liberté de choix de l'école apparaît comme un facteur négatif) ;*
- des moyens consacrés à des traitements élevés pour les enseignants plutôt qu'à de faibles tailles de classes.*

Ces conditions sont relatives : elles sont plus marquées dans certains des pays les plus performants, et elles connaissent des modalités qui peuvent être différentes selon les pays ; elles ne sont ni nécessaires, ni suffisantes, certains pays performants ne les connaissent pas, et certains pays qui les connaissent ne sont pas performants. Elles rejoignent cependant des conclusions déjà étayées dans la littérature internationale sur l'éducation. »

Il n'est donc pas question de copier. Comme toute autre évaluation, PISA ne saurait apporter automatiquement de solutions aux difficultés que rencontre notre système éducatif. Mais les recherches conduites et à conduire à partir de ses données, confrontées à d'autres, devraient permettre de débattre et de dégager des orientations.

Elles le pourraient, à mon sens, d'autant mieux qu' :

²⁰ PISA 2009 results, what makes a school successful? Volume IV, OCDE, 2010, PARIS. Il ne s'agit pas d'une citation, mais d'une traduction synthétique.

- on leur donnerait leur juste place dans l'ensemble des apports de la recherche à la décision :

« les enquêtes internationales comparatives des acquis des élèves constituent un « regard » extérieur qui doit impérativement être associé au « regard » intérieur apporté par les évaluations et études nationales. Elles doivent donc avoir toute leur place dans le dispositif d'évaluation et de pilotage de notre système éducatif, sans que cette place soit exclusive car, à défaut d'évaluations nationales directement liées aux objectifs propres de notre système éducatif, le risque pourrait exister de voir ériger les enquêtes internationales en modèle éducatif universel, voire, en caricaturant quelque peu, de faire de bons résultats à ces enquêtes la finalité principale ou unique du système éducatif. »²¹

- elles évolueraient vers plus de pertinence quant à l'éclairage des décisions :

« Les progrès les plus décisifs à accomplir aujourd'hui, dans la collecte de données organisée dans le cadre des cycles de PISA, ne sont ... sans doute pas du côté de la mesure des performances, mais bien plutôt du côté [d'une meilleure connaissance et d'une meilleure comparabilité] des conditions d'enseignement des élèves, de leur degré d'engagement dans les tâches qui leur sont proposées, etc., et surtout des caractéristiques « malléables » des systèmes éducatifs, c'est-à-dire celles sur lesquelles les responsables de ces systèmes peuvent agir : organisation des cursus, des procédures d'orientation, des modes d'évaluation, des processus et des niveaux de décision, etc. »²².

Il reste des marges à explorer autant chez les évaluateurs que chez les décideurs

Jean-Claude Émin.

²¹ *La France et les évaluations internationales des acquis des élèves*, Avis n°16 du Haut Conseil de l'évaluation de l'école, mai 2005.

²² Jean-Claude EMIN, *Que fait-on des évaluations internationales dans le système éducatif français ?* Colloque de l'ADMEE, Reims, octobre 2005, in « Evaluer pour former, Outils, dispositifs et acteurs », de Boeck université, Bruxelles, 2008.